Supporting Information

Zona et al. 10.1073/pnas.1516017113

SI Materials and Methods

Sites Description. Data for this study were collected from five EC towers across a 300-km latitudinal transect in the North Slope of Alaska. The three northernmost towers (BES, BEO, CMDL) are in the vicinity of Barrow, AK, where mean annual temperature is −11.3 °C and summer precipitation is 72 mm for the 1948 to 2013 period (37). The fourth site, ATQ, is about 100 km south from Barrow. Mean annual temperature and summer precipitation in ATQ are −10.8 °C and 100 mm, respectively, for the 1999 to 2006 period. The most southerly site (IVO) is located near the IVO Airstrip at the foothills of the Brooks Range Mountains, about 300 km south of Barrow, with a mean annual temperature and summer precipitation of −8.9 °C and 210 mm, respectively, for 2003 to 2008. The average snow depth was about 0.3 ± 0.1 m (mean ± SE) in BEO/BES, 0.2 ± 0.2 in ATQ, and 0.4 ± 0.1 in IVO. The vegetation is classified as W1 in Barrow [wet coastal plain dominated by sedges, grasses, and mosses (38)], as W2 in ATQ (tundra dominated by sedges, grasses, mosses, and some dwarf shrubs <40 cm tall), and as G4 in IVO [tussock-sedge dwarf-shrub, moss tundra (38)]. The land-cover types in Barrow and ATQ together are representative of about 60% of all arctic wetlands (38), whereas IVO represents the dominant vegetation type in Alaska. The zero curtain period (19) was defined as the fall period, when the soil temperature at 15- to 20-cm depth (the last soil layer to freeze in our system) was between 0.75 °C and −0.75 °C (Fig. S1 and Table S1). Among the Barrow sites, CMDL presented the least ice-wedge polygon development and is the driest. The presence of low center polygons results in the presence of wet waterlogged ponds interspersed with drier microtopographic features (high center polygons and/or polygons' rims) in the BEO site; BES is located in a vegetated drained lake with restricted drainage, the low topographic results in which being the wettest site.

Environmental Variables. A wide range of meteorological variables were measured at each of the five EC towers, including photosynthetic active radiation (PAR), which was measured with quantum sensors (LI-190; Li-COR) in all sites; net radiation was recorded using a net radiometer [REBS Q7 (Radiation & Energy Balance Systems, Inc.) in BES, BEO, CMDL, and ATQ; and an NR Lite (Kipp & Zonen) in IVO]; incoming solar radiation was measured using pyranometers (CMP3; Kipp & Zonen) in all sites; air temperature and relative humidity (RH) were measured with a Vaisala HMP 45 (in CMDL, BES, BEO, and ATQ), and a Vaisala 155a in IVO (and in BEO after 2013); soil heat flux at −2- to −5-cm depth was measured in four to six locations in all sites with REBS HFT3 (Radiation & Energy Balance); soil temperatures at different depths (at surface −1-, −5, −10-, −20-, and −30-cm depth in BES; at −5, −15, and −30 cm in four profiles in ATQ; and at the surface −5, −15, −30, and −40 cm in four profiles in IVO) were measured with thermocouples (either type-T or type-E; Omega Engineering); soil moisture was measured with a Water Content Reflectometer CS616 (Campbell Scientific) inserted perpendicularly (0–30 cm) or diagonally in different soil layers (0–10 cm and −20/−30 cm) in BES, or horizontally at different depths in the soil (−5, −15, and −30 cm in two different profiles in ATQ; and in three profiles in IVO). More details on these measurements can be found elsewhere (15, 18, 34). Snow depth was measured with Sonic Ranging Sensor in BEO/BES, ATQ, and IVO with an SR50A-L snow senor (Campbell Scientific).

Thaw depth and water table were measured manually during the summer and autumn on a weekly basis in the most accessible sites (BES, BEO, and CMDL) and once every 1–2 months in the more remote sites (ATQ and IVO). A graduated steel rod was used for the thaw-depth measurements, and PVC pipes with holes drilled every centimeter on their sides allowed for the water table measurements (15). Thaw-depth measurements were performed about every 5 m at 45 points in CMDL (three transects of 15 m each in the footprint of the EC tower); at 20 points in BEO, ATQ, and IVO (one transect in the footprint of each EC tower); and at 50 points every 4 m in the footprint of the EC tower at BES. Water table measurements were performed in these same plots for all sites with the exception of CMDL (where the installation of PVC was not allowed).

EC Data Processing and Data Filtering. Half-hourly fluxes were calculated using the EddyPro software (LI-COR), applying the following procedures and corrections: a despiking procedure of fast raw signals (39); the time lag between vertical wind velocity and scalar concentrations was computed based on the maximization of the covariance; turbulent departures from the means were calculated using linear detrending (40); a double-axis rotation and tilt correction was applied according to ref. 41; a correction of the high-pass filtering effect was applied (42); low pass-filtering effect resulting from instrumental attenuations was corrected using different procedures depending on the setup: the analytical method (43) was adopted for the open-path systems (LI-7700); the in situ spectral correction method (44) was used for the closed path (FGGA-24EP) analyzers, more suitable to describe site dependent spectral attenuation along the tubing system. For the closed path and enclosed analyzers (45), the compensation for density fluctuations was not needed as mixing ratio data were measured and used for flux computations (44). For the open-path LI-7700, a spectroscopic correction was computed (46). Spectral correction for instrument separation was applied (47); data quality control (QA/QC) flagging was computed based on stationarity and integral turbulence tests (48, 49), resulting in three flags (0, good; 1, intermediate; 2, poor); random uncertainty attributable to sampling errors was finally estimated following (50). One-point CH_4 storage term was computed based on the concentration measurements of the gas analyzer (51).

To prevent possible biases from the heating system of the METEK in IVO, we modified the activation of the heating of the sonic anemometer to only activate when the data quality was low (as indicated by a quality flag), instead of when temperature was below a set temperature threshold, as commercially available. From September 2014, the CSAT-3D were also externally heated using Freezstop Regular heating cables (Heat Trace), operated at 12 V direct current (DC). These insulated heating lines were cut to length to cover the support arms of the anemometers and yet were far enough removed from the transducer mounting arms to minimize flow distortion and other contamination of wind data. Control of the heating elements was done using the CR3000 data logger and a normally closed solid-state DC relay. The data logger program activated a relay when the sonic transducers were blocked by snow and/or ice, as reported by the diagnostic output by the CSAT-3D. All data when the heating was active were removed for the heated anemometer. Additional data cleaning was performed accordingly to the following criteria: data were removed when the quality flags (48, 49) of H, LE, and CH_4 fluxes were 2; when the internal pressure of the LGR was

≤132 torr (corresponding to malfunctioning of the instrument); or when identified as an outlier (when exceeding a moving-window weekly 1 SD of the individual fluxes).

We used Los Gatos Research FGGA gas analyzers at all sites, except in IVO, where low power availability restricted use to the open-path LI-7700 analyzer. A second LI-7700 was also implemented in CMDL alongside the FGGA to ensure comparability of the results using these two instruments. To this end, in CMDL, we cross-compared the CH_4 fluxes estimated using these two instruments from October 1, 2013 to October 1, 2014 (Fig. S5). The difference between the LICOR and LGR-FGGA-24EP CH_4 fluxes was calculated directly and the resulting distribution used to elucidate features of their uncertainty (Fig. S5). Half-hourly fluxes were used to provide the largest possible sample size. The mean difference (LICOR F_{CH4} – LGR-FGGA-24EP F_{CH4}; ΔF_{CH4}) was −0.0226 mg m^{-2} h^{-1}. Considering that the overall mean values of the LICOR and LGR-FGGA-24EP fluxes were 0.216 and 0.239 mg m^{-2} h^{-1}, respectively, this amounts to about a 10% difference between the two sensors, with the average LICOR flux slightly lower than the LGR-FGGA-24EP. The data were heteroscedastic, whereby the uncertainty estimated from Laplace distributions (Fig. S5) of data binned by flux magnitude increased with the value of the fluxes (52, 53). The average annual data coverage of the CH_4 fluxes (after removal of the data as indicated above) for the entire measuring period was 52% (CMDL LGR), 29% (CMDL LI-7700), 54% (BEO LGR), 58% (BES LGR), 49% (ATQ), and 38% (IVO LI-7700). The seasonal distribution of the data coverage is shown in Fig. S4.

Gap Filling of the Eddy Covariance CH_4 Fluxes. Missing CH_4 flux data were gap-filled by applying artificial neural networks as described in refs. 54 and 55. The following meteorological variables, that most commonly act as drivers for CH_4 emissions, were used following the principle of parsimony and avoiding unnecessary input variables of a cross-correlative and cross-dependent nature: air temperature, air pressure, solar radiation, vapor pressure deficit, soil temperature, soil moisture, and the decomposed wind speed and direction, together with the fuzzy sets representing the annual seasons, as discussed by Dengel et al. (55). A training and a testing dataset for the neural networks were produced, including both daytime and nighttime periods in an equal manner covering all meteorological conditions from both years.

Several combinations of input variables and neurons were tested before we decided to include the same input variables and eight neurons for all datasets to keep the method uniform across all five sites. Each site analysis included 500 repetitions of which the 25 best [correlation coefficient values of the predicted (testing stage) values] runs were included in the gap filling itself. The application of error analysis, the agreement or disagreement between measured and modeled data helped to estimate the overall performance of the neural network models. The overall performance of all five models show no under- or overestimation of the true fluxes. Fig. S6 presents the gap-filled data (red) and the original measurements (gray), and Fig. S7 presents the influence of the gap filling on the daily averaged CH_4 fluxes. Final error analysis results can be found in Table S2, including the correlation of determination (R^2), together with the root mean square error (RMSE) converted to true physical units of mg C-CH_4 m^{-2} h^{-1} indicating the uncertainties of the models.

Circumpolar Arctic CH_4 Emissions Estimates. For their central estimate in the CH_4 emission of Arctic tundra, McGuire et al. (8) used estimated wetland areas of 772,076; 7,540; 18,139; and 812,969 km^2 for North America, North Atlantic, Northern Europe, and Eurasia subregions, respectively, for a total of 1,610,724 km^2; whereas we used the Circumpolar Arctic Vegetation map and included the land cover types B3, B4, G1, G2, G3, G4, P1, P2, S1, S2, W1, W2, and W3, so encompassing all tundra types excluding glaciers and lakes (38), which resulted in a total area of 5,070,000 km^2 across the entire Arctic. Our choice was justified by the importance of upland tundra (e.g., ATQ, CMDL, IVO). The land area we used was greater, but the measured fluxes used were lower than those used by McGuire et al. (8), who included data from boreal wetlands in Alaska (20, 56) in estimating rates of CH_4 emissions. Fortuitously, the differences in land area and fluxes compensated, resulting in similar annual estimates of CH_4 flux to those reported here.

Regional Fluxes (CARVE). The aircraft fluxes were at times lower than the mean of the EC tower fluxes, as has been observed previously in point-scale and regional-scale flux comparisons. The influence of the Brooks Mountain Range has been excluded (Fig. S3 *C* and *F*, gray shading). Additionally, we note that WRF estimates of planetary boundary layer (PBL) ventilation rates are difficult to assess quantitatively and might be subject to particular bias in the fall, when heat fluxes are low. A ∼28% difference of CO_2 eddy fluxes from aircraft compared with towers has been reported (57) and attributed in part to differences in the aircraft and tower footprints. As seen in Fig. S3, the footprint of the regional flux is much larger than that of the flux towers and includes areas assumed to be less productive particularly in autumn and winter, such as frozen lakes. Nevertheless, the regional CH_4 fluxes strongly support the view that our EC fluxes capture relevant cold season CH_4 dynamics and the response of CH_4 emissions to soil climate across the wider North Slope area.

Regional fluxes of CH_4 were estimated with aircraft data from the CARVE 2012 to 2014 (30). CH_4-, CO_2-, and CO-mixing ratios were measured using two independent cavity ring-down spectrometers: one operated wet (58) (G1301-m; Picarro) and one dry (30) (G2401-m; Picarro). Each analyzer was calibrated throughout the flight, ensuring a continuous 5-s time series. Ozone-mixing ratios were measured using a 2B Technologies model 205. The aircraft data were aggregated horizontally every 5 km and vertically in 50-m intervals below 1 km and 100-m intervals above 1 km. Each aggregated position was treated as a receptor in a Lagrangian transport model (WRF-STILT), which calculated the back trajectory of 500 particles from each receptor location. WRF-STILT represents the Stochastic Time-Inverted Lagrangian Transport (STILT) model coupled with meteorology fields from the polar variant of the Weather and Research Forecasting model (WRF) [v3.5.1 (59)]. The WRF-STILT calculation allowed for the quantification of the space and time where upstream surface fluxes influenced the measured mixing ratios. A total 24 h 2D surface influence field (i.e., footprint) was calculated for each flight (e.g., Fig. S3 *C* and *F*), representing the response of the receptor to a unit surface emission (ppb/mg C-CH_4 m^{-2} h^{-1}) of CH_4 in each grid square (0.5° × 0.5° grid). The systematic uncertainty of the calculated surface influence is estimated at 10–20% (59). For comparison with the flux towers, aircraft data were carefully selected and a number of assumptions made to calculate a regional flux of CH_4. Only data collected north of 68° N, west of −153° W, with CO <150 ppb (to remove impacts of anthropogenic influence in the Deadhorse/Prudhoe Bay area), below 1,500 m above ground level and with over 60% surface influence from the North Slope were selected. CH_4 emissions from the higher altitudes of the Brooks Mountain Range were assumed to be negligible (gray area in Fig. S3 *C* and *F*). Assuming a uniform land surface emission, CH_4-mixing ratios should scale linearly with the total land surface influence observed at that receptor point. The flux of CH_4 for each flight day was calculated from the correlation of CH_4 with the STILT land surface influence [ordinary least-square (OLS) regression; Fig. S3 *A* and *D*], where the slope represents the regional flux and the intercept is the regional background CH_4-mixing ratio, which was assumed not to vary greatly during the flight. The

ozone deposition velocity was also calculated in a similar manner (Fig. S3 *B* and *E*), and only flights with an ozone deposition velocity consistent with the expected seasonal cycle were used as a valid CH_4 flux.

Statistical Analysis. To understand the environmental control on CH_4 fluxes over the year, weekly median CH_4 fluxes were modeled as a function of the weekly averaged environmental data. Environmental variables included were soil temperature at 0–5, 10–15, and 20–30 cm; soil water content (SWC) at 0–10 and 20–30 cm; and thaw depth. Only weeks with more than 20 flux and environmental measurements were included in the analysis. Because thaw-depth measurements were collected once a week in the Barrow sites, the weekly average of all available flux and environmental measurements were used in the statistical analyses. The remoteness of the IVO and ATQ sites limited the frequency of visits and therefore of thaw depth measurements. As a result, values were linearly interpolated between measurements during the summer. Because we measured a consistent thaw depth from the end of the summer until October 2014, we assumed a stable depth of the active layer from the end of summer throughout the zero curtain period, and we used the soil temperature profiles and the duration of the zero curtain to extrapolate the thaw depth until soil freezing. After soil freezing, we set the thaw depth to zero until the beginning of the following spring (again defined by the soil temperature profiles, as defined in Table S1). Because CH_4 fluxes presented a skewed distribution, they were log-transformed for all of the statistical analysis. Because of the data loss in the soil environmental data in 2013 and 2014 (e.g., the soil moisture data from IVO and soil temperature data in ATQ), we extended the dataset used for the statistical analysis to include data until May 2015, covering a full year for each site. Only the three sites with the most complete environmental datasets were used in the statistical analysis (ATQ, IVO, and BES).

Linear and nonlinear mixed effects models were used for this analysis (using the lme4 and the nlme in R; R Developing Core Team). The mixed-effects models included the "week" of measurement and "site" as continuous and categorical random effects, respectively, to account for the pseudoreplication and the different sites measured. The nonlinear mixed-effects model was used to test whether an exponential or power fit of soil temperature, and SWC were the best predictor of CH_4 fluxes. However, the increase in the complexity of the model did not justify the use of nonlinear mixed effects modeling, as shown by Akaike information criterion (AIC) values and the partial *F* test of the two models. Therefore, the results of the linear mixed effects models are reported here. The model performance was evaluated based on the AIC values, on the significance of the partial *F* test (used to compare two models), and on the marginal coefficient of determination (similar to the explanatory power of the linear models) for generalized mixed-effects models as output by the r.squaredGLMM function within the MuMIn package in R (60, 61).

The best univariate model explaining the variability in the CH_4 fluxes during the entire year was soil temperature (T) at 20–30 cm, with an AIC of 5.5 and a marginal coefficient of determination of 0.85. During the cold period (from September to May), soil water content at 20- to 30-cm depth was the most important variable explaining CH_4 fluxes, with an AIC of −27 and a marginal coefficient of determination of 0.89. During the summer period, instead, the best univariate model included thaw depth, presenting an AIC of 1.7 and a marginal coefficient of determination of 0.89. The best multivariate model for the CH_4 fluxes during the entire year included soil water content at 20–30 cm in addition to soil T at 20–30 cm, with an AIC of −23 and a marginal coefficient of determination of 0.89. This multivariate model was significantly different in its explanatory power from the univariate model that only included soil T at 20–30 cm (as shown by the significant partial *F* test of the difference in the two models). Similarly, during the cold season, the best multivariate model included soil T at 20–30 cm and soil water content at 20–30 cm, with an AIC of −55 and a coefficient of determination of 0.89 (which was significantly different from the univariate model as shown by the partial *F* test). No multivariate model was statistically different from the univariate one (which only included thaw depth) in the summer.

Surface Inundation at the Tower Sites. Seasonal patterns of aerial proportion (%) of surface water inundation within 25 × 25 km footprints extending over the greater Alaska domain is derived from K-band passive microwave satellite remote-sensing (62, 63). Surface inundation at the tower sites persists from late May following surface thaw through November, when colder air temperatures minimize the presence of liquid water above the soil surface. Wet subsurface soil conditions at ATQ and IVO contribute to peak summer CH_4 emissions of 20–50 mg C-CH_4 m^{-2} d^{-1}, despite lower surface water inundation relative to the CMDL, BES, and BEO tower sites, indicating that substantial CH_4 emissions are not confined to wet, inundated tundra.

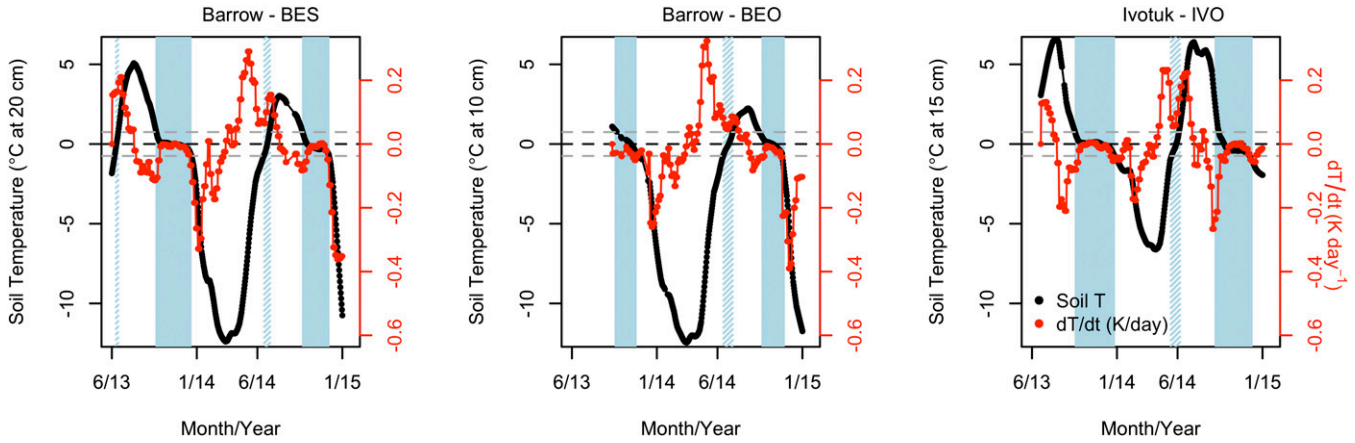

Fig. S1. Zero curtain assessment. Soil temperatures at BES (20 cm, left), BEO (10 cm, center), and IVO (15 cm, right), recorded during 2013 and 2014. The red curves show the time derivative of temperature (°C/d), illustrating the effect of latent heat in stabilizing the temperature when the soils are poised in the zero curtain (0 ± 0.75 °C, light blue shading and horizontal dashed lines). Temperature data have been smoothed using a locally weighted least-squares linear trend ("lowess" filter) with a window of 35 d, roughly equivalent to a two-sided moving average of ±1 wk.

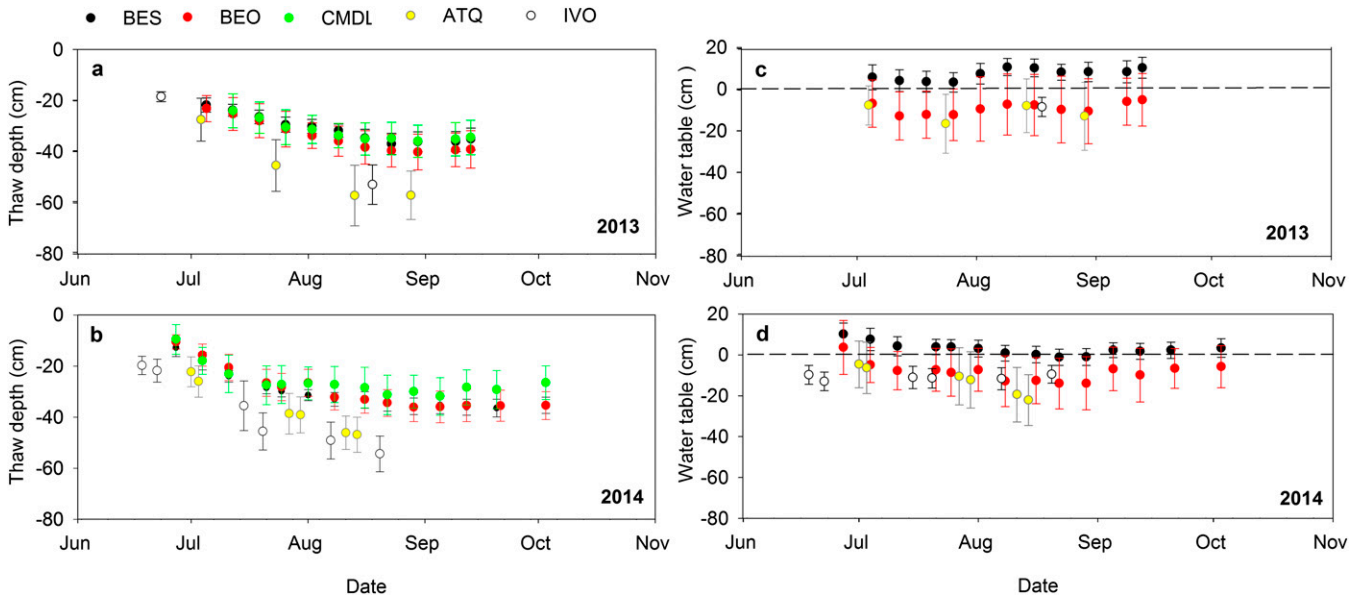

Fig. S2. Thaw depth and water table. Active layer depth (thaw depth) in centimeters from surface (*A* and *B*) and water table (*C* and *D*) depths at four of the five sites during summer and fall 2013 and 2014 (negative values indicate thaw depth and water table below the surface). PVC pipes for water table monitoring were not allowed at the Barrow CMDL site because of clean air restriction by the NOAA. Displayed are averages and SEs from the mean (BEO, ATQ, and IVO: n= 20; CMDL: n= 45; BES: n= 50).

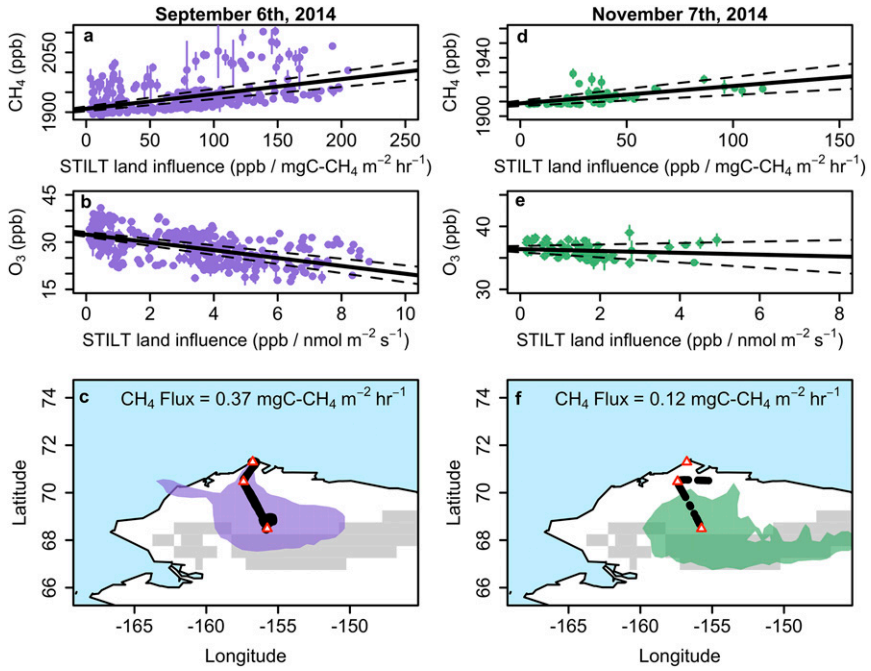

Fig. S3. Methane and ozone aircraft data from CARVE during 2 d in 2014: September 6 and November 7. (*A* and *D*) Correlation of observed CH_4 in ppb on the *y* axis, with the total land influence from the STILT footprint for the matching receptor location on the *x* axis. (*B* and *E*) Correlation of observed O_3 with this same total STILT influence (64). Solid lines show the ordinary least-squares linear regression of these relationships, where the slope of each regression is the calculated flux. (*C* and *F*) Composite footprint for the data used to calculate the flux for each day. Black lines indicate the flight track, and the white triangles with red edges indicate the locations of the eddy flux sites. Gray shading indicates the mountains of the Brooks Range, where fluxes were assumed to be negligible.

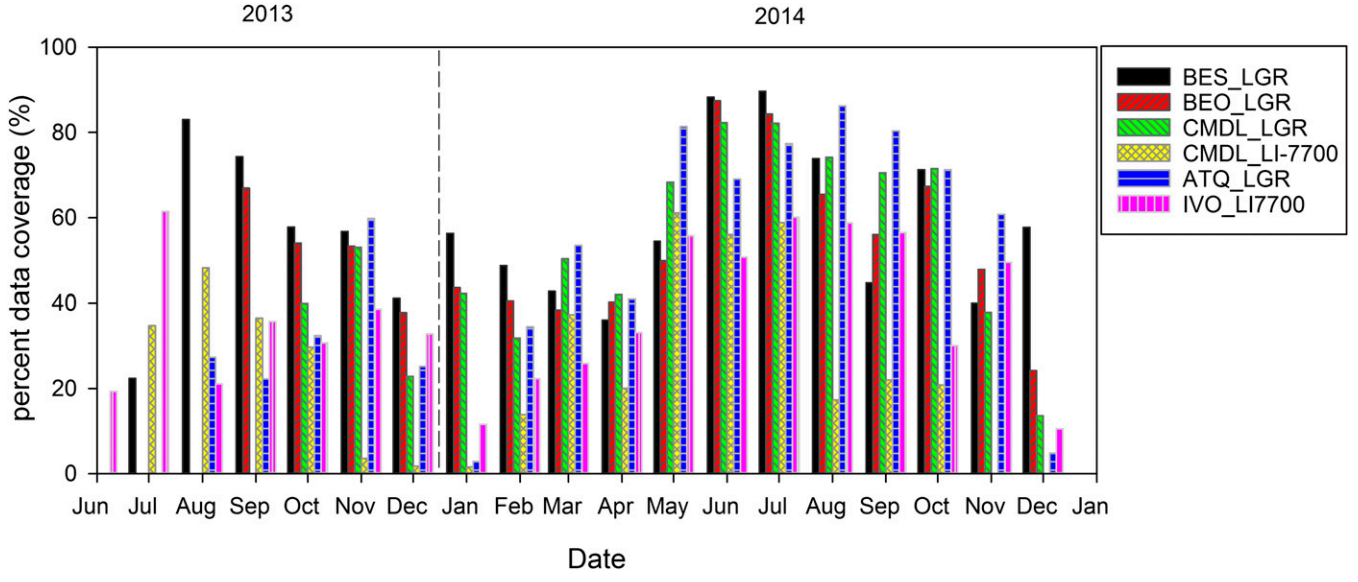

Fig. S4. Percentage of data coverage. Percentage of data coverage at each of the sites of the EC CH_4 fluxes data used in this study.

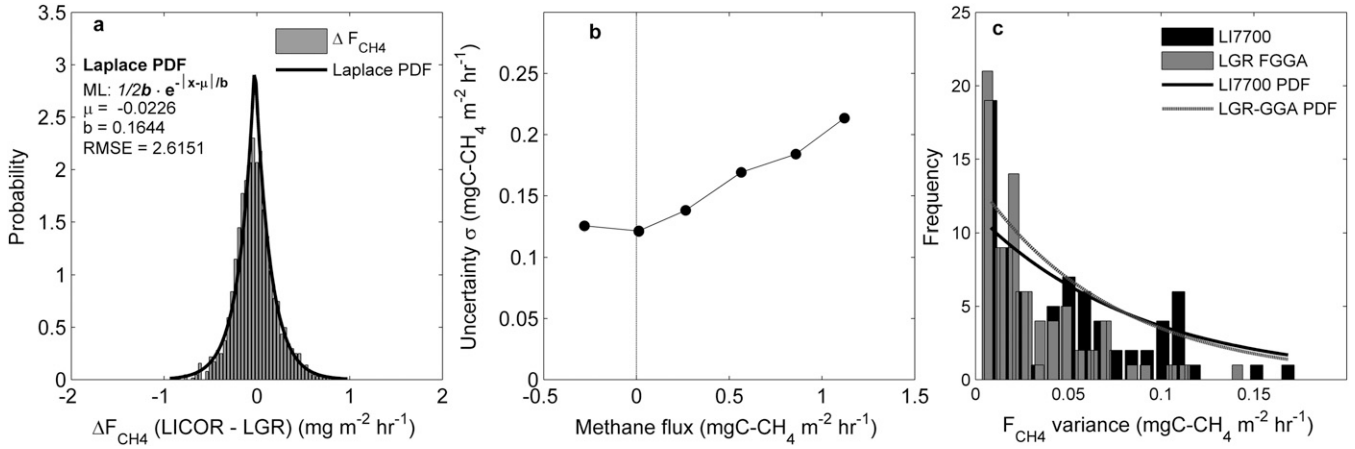

Fig. S5. Comparison of CH_4 fluxes from the LGR-FGGA-24EP and LI-7700 in Barrow (CMDL). (*A*) Differences in CH_4 fluxes derived from the LICOR 7700 and LGR-FGGA-24EP analyzers at CMDL from October 1, 2013 to October 1, 2014 represented by a Laplace probability density function (PDF) (n = 1,681). The histogram and PDF are normalized by trapezoidal integration. The ΔFCH_4 values were calculated as LI-7700 − LGR FGGA-24EP CH_4 fluxes. (*B*) Uncertainty in FCH_4 was binned by flux magnitude. The σ values were calculated as the Laplace analog to SD ($\sqrt{2}$ b). The uncertainty associated with LI-7700– and LGR-FGGA-24EP–derived fluxes increases with the magnitude of the flux (i.e., heteroscedasticity). (*C*) Daily variance in CH_4 fluxes calculated from the LI-7700 and LGR-FGGA-24EP. Fitted lines are exponential PDFs with an estimated μ of 0.087 (95% CI: 0.06–0.14) mg C-CH_4 m^{-2} h^{-1} for the LI-7700 and 0.073 (95% CI: 0.05–0.12) mg C-CH_4 m^{-2} h^{-1} for the LGR-FGGA-24EP.

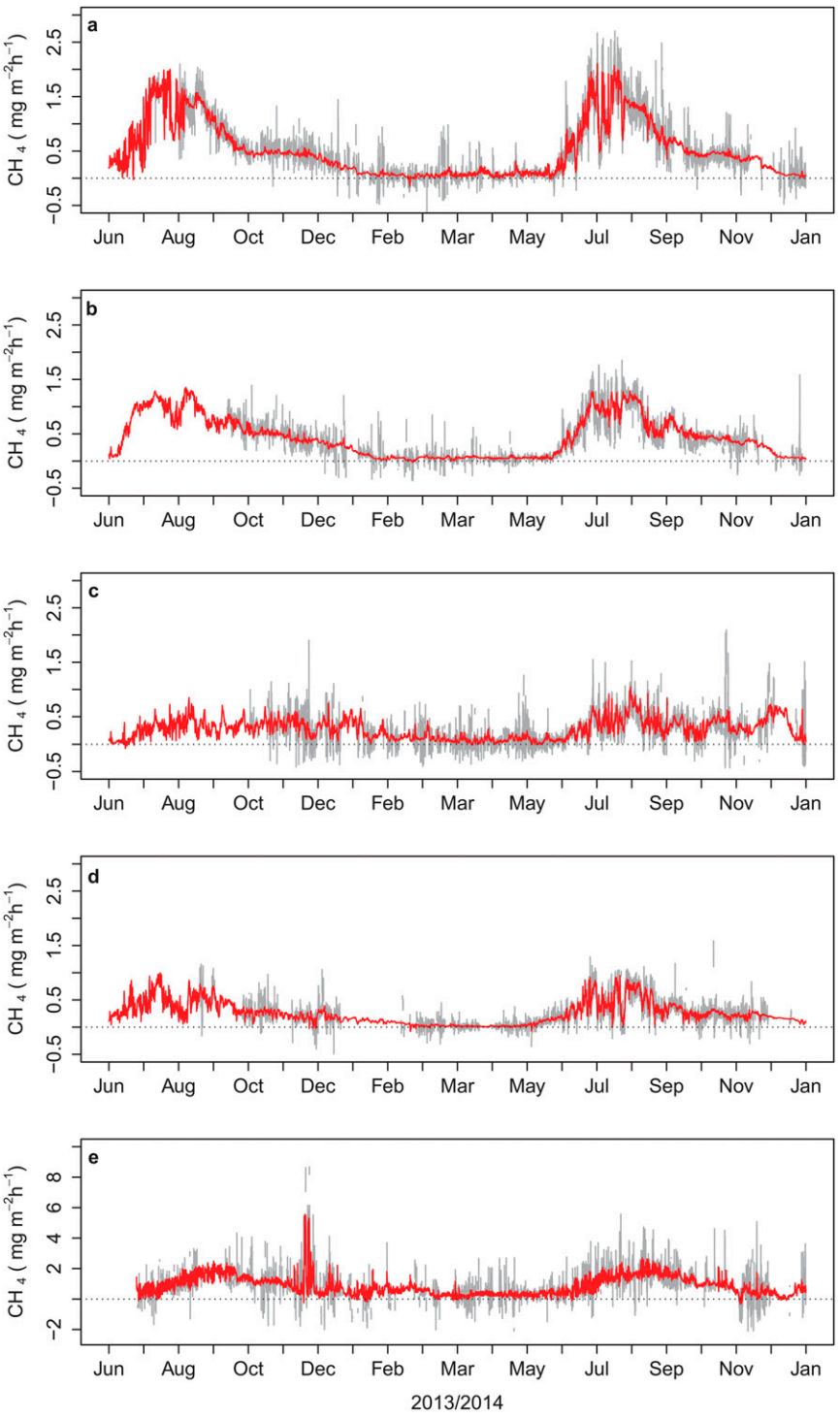

Fig. S6. Gap filling of the methane fluxes. Gap-filled CH_4 fluxes (red) superimposed on the measured fluxes (in gray) for the indicated sites: BES (*A*), BEO (*B*), CMDL (*C*), ATQ (*D*), and IVO (*E*).

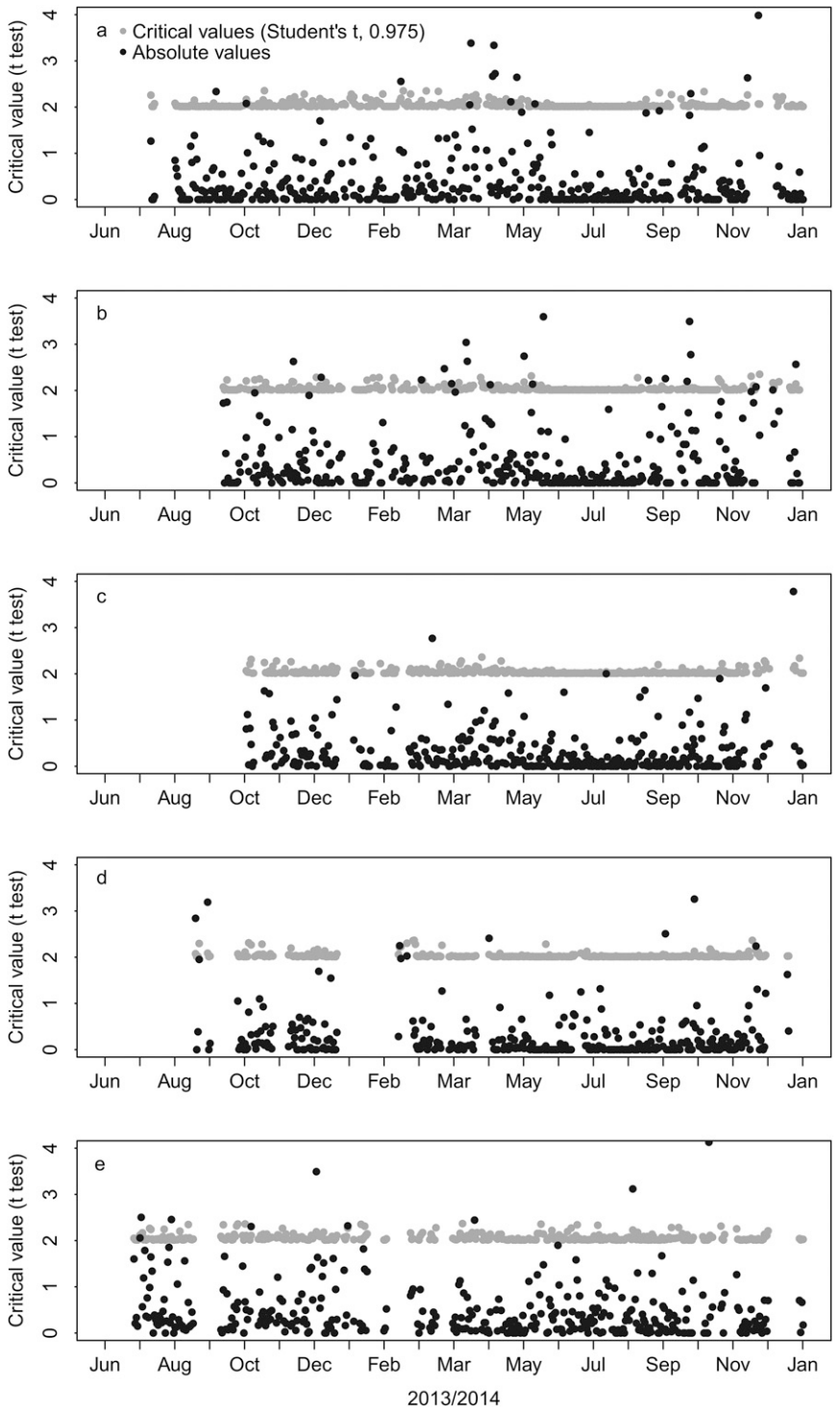

Fig. S7. Critical values of the Student's *t* test distribution of the gap filling for all of the sites. We applied the critical values of the Student's *t* test to each day where more than 6 h of data were available to investigate the influence of the gap filling on the daily averages of CH_4 fluxes: BES (*A*), BEO (*B*), CMDL (*C*), ATQ (*D*), and IVO (*E*). Student's *t* test values higher (black full circles) than the critical values of the Student's *t* distribution (0.975 level) (gray full circles) represent days on which the gap filling of the data had an impact on the final daily average.

Table S1. Zero curtain and spring thawing estimates in 2013 and 2014

Year and site	Spring start	Spring end	Spring thawing days	Autumn start	Autumn end	Zero curtain days	Soil T depth, cm
2013							
BES	—	—	—	261	352	91	20
ATQ	—	—	—	260	343	83	15
BEO	—	—	—	259	314	55	10
IVO	—	—	—	259	360	101	15
CMDL	—	—	—	—	—	—	—
2014							
BES	169	184	15	264	333	69	20
ATQ	154*	174*	20*	264	319	55	15
BEO	167	190	23	262	321	59	10
IVO	135	158	23	245	340	95	15
CMDL	—	—	—	—	—	—	—

Zero curtain dates in the autumn and spring thawing dates and duration are determined from the criteria $-0.75 < T < 0.75$ and are shown in Fig. S1. Because of data loss in the soil temperature data, an average of BES and BEO was used for CMDL. Because spring 2013 was not consistently collected in all sites, we estimated the zero curtain periods starting from autumn 2013. Dashes indicate that no data is available.
*ATQ spring 2013 and autumn 2014.

Table S2. Error analysis of the gap filling of the methane fluxes

Values	BES	BEO	CMDL	ATQ	IVO
Number data points, *n*	8,350	6,978	6,891	6,930	6,309
Data coverage, %	60.0	50.0	49.5	49.8	45.4
Neurons	8	8	8	8	8
Runs	100/500	100/500	100/500	100/500	100/500
Final averaged runs	25	25	25	25	25
Daytime training data coverage, %	49	51	50	49	51
Nighttime training data coverage, %	50	48	49	50	48
Daytime testing data coverage, %	49	50	50	51	49
Nighttime testing data coverage, %	50	49	49	48	50
R^2-training	0.91	0.88	0.61	0.76	0.57
R^2-testing	0.89	0.83	0.52	0.75	0.45
Mean original, true units	0.56	0.41	0.26	0.25	0.89
Mean modeled, true units	0.51	0.43	0.27	0.25	0.86
Mean gap-filled, true units	0.51	0.43	0.27	0.25	0.86
Mean RMSE, true units	0.171	0.15	0.181	0.125	0.69

Error analysis results of the gap filling of the CH_4 fluxes for each of the indicated sites, including the coefficient of determination (R^2) and the RMSE converted to true physical units of mg C-CH_4 m^{-2} h^{-1} indicating the uncertainties of the models.